Výpočtové učenie - 12. prednáška

Gabriela Andrejková
Ústav informatiky, Prírodovedecká fakulta, UPJŠ, Košice

May 2, 2004

1 Zložitosť vzorky konzistentných algoritmov

Fakt: Ak hypotézový priestor H má konečnú VC dimenziu, tak je potencionálne naučiteľný. Inak povedané:

K ľubovoľným parametrom δ - dôveryhodnosť, ϵ - presnosť $(0 < \delta, \epsilon < 1)$, existuje vzorka dĺžky $m_0 = m_0(H, \delta, \epsilon)$ taká, že

$$m \geq m_0 \Longrightarrow \mu^m\{s \in S(m,t); H[s] \cap B_\epsilon = 0\} > 1 - \delta \tag{1}$$

pre ľubovoľné pravdepodobnostné rozdelenie μ na X ľubovoľný cieľový koncept $t \in H$.

Teda z toho vyplýva, že

- **ľubovoľný konzistentný učiaci algoritmus L pre H je *pac* a ďalej**
- **ľubovoľné $\mu_0(H, \delta, \epsilon)$, pre ktoré platí vyššie uvedená podmienka, je horná hranica na zložitosť vzorky $m_L(H, \delta, \epsilon)$.**

Ukážeme, že existuje presnejší výraz pre m_0, a teda pre hornú hranicu na zložitosť vzorky ľubovoľného konzistentného učiaceho algoritmu L pre H.

Pripomeňme, že bolo dokázané, že ak H je konečný hypotézový priestor a L je konzistentný učiaci algoritmus pre H, tak L je *pac* a platí

$$m_L(H, \delta, \epsilon) \leq \left\lceil \frac{1}{\epsilon} ln(\frac{|H|}{\delta}) \right\rceil \tag{2}$$

Horná hranica, ktorú odvodíme, závisí od VC dimenzie H a nie od mohutnosti H.

Veta: Predpokladajme, že H je hypotézový priestor s konečnou VC dimenziou $d \geq 1$ a platí $0 < \delta, \epsilon < 1$. Nech

$$m_0 = m_0(H, \delta, \epsilon) = \left\lceil \frac{4}{\epsilon}(d.lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta})) \right\rceil \tag{3}$$

Potom pre ľubovoľné $m \geq m_0$ platí

$$\mu^m\{s \in S(m,t); H[s] \cap B_\epsilon \neq \emptyset\} < \delta. \tag{4}$$

Dôsledok: Predpokladajme,že hypotézový priestor H má VC dimenziu $d \geq 1$. Potom ľubovoľný konzistentný učiaci algoritmus L pre H je *pac* so zložitosťou vzorky

$$m_L(H, \delta, \epsilon) \leq \left\lceil \frac{4}{\epsilon}(d.lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta})) \right\rceil \tag{5}$$

Toto už je sľúbený výsledok.

Príklad 1: Nech H je priestor lúčov. Jeho VC dimenzia je $VCdim(H) = 1$, teda ak L je ľubovoľný konzistentný učiaci algoritmus pre priestor lúčov, máme

$$m_L(H, \delta, \epsilon) \leq \left\lceil \frac{4}{\epsilon}(d.lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta})) \right\rceil \tag{6}$$

My sme priamo dokázali, že

$$m_0 = \left\lceil \frac{1}{\epsilon} ln(\frac{1}{\delta}) \right\rceil. \tag{7}$$

Vidíme, že hranica vypočítaná priamo je lepšia než vyplývajúca z VC dimenzie. Avšak priame argumenty sú mnohokrát ťažké a v tomto prípade je zrejmé, že ak δ a ϵ sú toho istého radu, tak tieto hranice sa líšia len o konštantu.

Príklad 2: Reálny perceptrón P_n má VC dimenziu $n+1$. Predpokladajme, že pre ľubovoľnú tréningovú vzorku pre hypotézu v P_n máme nájsť stav ω perceptrónu taký, že h_ω je konzistentná so vzorkou. Potom keď použijeme tréningovú vzorku dĺžky

$$\left\lceil \frac{4}{\epsilon}((n+1).lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta})) \right\rceil \tag{8}$$

máme zaručenú približne správnu hypotézu bez ohľadu na cieľovú hypotézu a pravdepodobnostné rozdelenie príkladov.

~~2~~ 1 Dolné hranice pre zložitosť vzorky

Pripomeňme, že sme dokázali, že ak priestor H má nekonečnú VC dimenziu, tak [illegible] nie je potenciálne naučiteľný.

Veta: Predpokladajme, že hypotézový priestor H má VC dimenziu $d \geq 1$. Potom existuje konzistentný *pac* učiaci algoritmus L pre H taký, že pre ľubovoľné δ a ϵ zložitosť vzorky splňuje

$$m_L(H, \delta, \epsilon) \geq d(1 - \epsilon).$$

Uvedený výsledok je síce jednoduchý, ale je použiteľný len na konzistentné učiace algoritmy. Ďalej neposkytuje univerzálnu dolnú hranicu na zložitosť vzorky konzistentného učiaceho algoritmu; skôr sa zaoberá najhoršími možnými konzistentnými učiacimi algoritmami. Ukážeme silnejší výsledok Ehrenfeuchta et al. - 1989, ktorý poskytuje dolnú hranicu na zložitosť vzorky *ľubovoľného pac* učiaceho algoritmu pre hypotézový priestor konečnej VC dimenzie.

Veta: Pre ľubovoľný hypotézový priestor H s VC dimenziou $d \geq 1$ a pre ľubovoľný *pac* učiaci algoritmus L pre H platí

$$m_L(H, \delta, \epsilon) > \frac{d-1}{32\epsilon} \tag{9}$$

pre $\delta \leq \frac{1}{100}$ a $\epsilon \leq \frac{1}{8}$.

Ak by sme uvažovali, že zložitosť vzorky prekročí $(d_0 - 1)/32\epsilon$, kde d_0 je ľubovoľné kladné číslo spĺňajúce $VCdim(H) \geq d_0$, tak by sme mohli dokázať nasledujúce tvrdenie:

Tvrdenie: Ak hypotézový priestor H má konečnú VC dimenziu, tak neexistuje žiadny *pac* učiaci algoritmus pre H.

Tieto výsledky podporujú tvrdenie, že VC dimenzia je dobrá miera "expresívnej sily" hypotézového priestoru H: väčšia VC dimenzia H znamená väčšiu zložitosť vzorky pre *pac* učenie H. V skutočnosti výsledky môžu byť zovšeobecnené aj pre prípady, keď C je ľubovoľný konceptový priestor s VC dimenziou aspoň $d_0 \geq 1$ a H je ľubovoľný hypotézový priestor (nie nutne rovný C).

Ak L je učiaci algoritmus pre (C, H), vstupom do L musí byť tréningová vzorka dĺžky väčšej než $(d_0 - 1)/32\epsilon$, aby bola zaručená presnosť $\epsilon \leq \frac{1}{8}$ s pravdepodobnosťou $1 - \delta > 99/100$. Ak C má nekonečnú VC dimenziu, tak nemôže existovať žiadny učiaci algoritmus pre (C, H), ktorý je *pac* pre ľubovoľný hypotézový priestor H.

Príklad: Ak J je priestor všetkých zjednotení intervalov, tak pretože J má nekonečnú VC dimenziu, neexistuje žiadny *pac* učiaci algoritmus pre (J, H) pre ľubovoľný hypotézový priestor H. Vyššie uvedený výsledok je veľmi silný. Ukazuje nielen to, že neexistuje žiadny konzistentný alebo efektívny *pac* učiaci algoritmus, ale tiež, že pre dané neohraničené výpočtové zdroje žiadny algoritmus nemôže *pac* naučiť J nezávisle od toho ako by boli reprezentované výstupné hypotézy. Samozrejme, tieto závery platia pre ľubovoľný priestor nekonečnej VC dimenzie, takých ako priestor všetkých uzavretých množín alebo priestor charakteristických funkcií všetkých polygonálnych oblastí v R^2.

Iný výsledok týkajúci sa dolných hraníc zaviedol Blumer et al., 1989. Tento výsledok obsahuje δ a ϵ, ale nezávisí od VC dimenzie hypotézového priestoru. Toto sa dá použiť na netriviálne hypotézové priestory, t. j. také priestory, ktoré obsahujú viac než dve hypotézy.

Veta: Predpokladajme, že L je ľubovoľný *pac* učiaci algoritmus pre netriviálny hypotézový priestor H. Potom

$$m_L(H, \delta, \epsilon) > \frac{(1-\epsilon)}{\epsilon} ln(\frac{1}{\delta}), \tag{10}$$

pre ľubovoľné $0 < \delta, \epsilon < 1$.

3 Porovnanie hraníc zložitosti vzoriek

Ako už bolo uvedené, mnohé predchádzajúce výsledky môžu byť zovšeobecnené pre prípad, keď konceptový a hypotézový priestor sú rôzne. Ukážeme hranice pre zložitosť vzorky v týchto všeobecnejších prípadoch.

Veta: Nech C je konceptový a H je hypotézový priestor a predpokladajme, že H má konečnú VC dimenziu aspoň 1. Ak L je ľubovoľný konzistentný učiaci algoritmus pre (C, H), potom L je *pac* a pre zložitosť vzorky platí

$$m_L \leq \left\lceil \frac{4}{\epsilon}(VCdim(H) lg(\frac{12}{\epsilon}) + lg(\frac{2}{\delta}))\right\rceil \tag{11}$$

pre ľubovoľné δ a ϵ.

Veta: Nech C je konceptový a H je hypotézový priestor taký, že C má VC dimenziu aspoň 1. Predpokladajme, že L je ľubovoľný *pac* učiaci algoritmus pre (C, H). Potom pre zložitosť vzorky pre L platí

$$m_L > max(\frac{VCdim(C) - 1}{32\epsilon}, \frac{1}{\epsilon} ln(\frac{1}{\delta})) \tag{12}$$

pre všetky $\delta \leq 1/100$ a $\epsilon \leq 1/8$.

Označenie:

Píšeme $f = O(g)$, ak existuje nejaká konštanta $k > 0$ taká, že pre všetky relevantné x platí $f(x) \leq k.g(x)$.

Píšeme $f = \Omega(g)$, ak existuje nejaká konštanta $k > 0$ taká, že pre všetky relevantné x platí $f(x) \geq k.g(x)$.

Použitím toho označenia sformulujeme vzťahy pre zložitosť vzorky.

- Ak L je *pac*, potom C musí mať konečnú VC dimenziu a platí

$$m_L(C, \delta, \epsilon) = \Omega(\frac{VCdim(C)}{\epsilon} + \frac{1}{\epsilon} ln(\frac{1}{\delta})) \tag{13}$$

- Ak H má konečnú VC dimenziu a L je konzistentný, potom L je *pac* a platí

$$m_L(C, \delta, \epsilon) = O\Big(\frac{VCdim(C)}{\epsilon} ln(\frac{1}{\epsilon}) + \frac{1}{\epsilon} ln(\frac{1}{\delta})\Big) \tag{14}$$

- Ak H je konečný a L je konzistentný, potom L je *pac* a platí

$$m_L(C, \delta, \epsilon) = O\Big(\frac{1}{\epsilon} ln|H| + \frac{1}{\epsilon} ln(\frac{1}{\delta})\Big) \tag{15}$$

V prípade, že $C = H$, VC dimenzia d je konečná a L je konzistentný, máme dolné a horné hranice

$$m_L(H, \delta, \epsilon) = \Omega\Big(\frac{d}{\epsilon} + \frac{1}{\epsilon} ln(\frac{1}{\delta})\Big);$$

$$m_L(H, \delta, \epsilon) = \emptyset\Big(\frac{d}{\epsilon} ln(\frac{1}{\epsilon}) + \frac{1}{\epsilon} ln(\frac{1}{\delta})\Big);$$

Faktor $ln(1/\epsilon)$, ktorý odlišuje hornú hranicu od dolnej, je nevyhnutný. Výsledky Hausslera, Littlestonea a Warmutha, 1988, ukazujú, že pre každé $d \geq 1$ existuje hypotézový priestor H_d a konzistentný učiaci algoritmus L pre H_d so zložitosťou vzorky rovnou hornej hranici. Na druhej strane je otvorený problém rozhodnúť, či pre každé d a pre každý konceptový priestor C s VC dimenziou d, existuje *nejaký* hypotézový priestor H a *nejaký* (C, H) učiaci algoritmus L, pre ktorý zložitosť vzorky má dolnú hranicu.

Cvičenia: